新疆野苹果叶绿体基因组特征及其进化历史

张 健[1,2]， 张宏祥[1,3,4]

（1. 中国科学院新疆生态与地理研究所，荒漠与绿洲生态国家重点实验室，干旱区生态安全与可持续发展重点实验室，新疆 乌鲁木齐 830011；2. 中国科学院大学，北京 100093；3. 新疆抗逆植物基因资源保育与利用重点实验室，新疆 乌鲁木齐 830011；4. 中国科学院新疆生态与地理研究所标本馆，新疆 乌鲁木齐 830011）

摘 要：新疆野苹果是苹果属重要的种质资源，属国家二级保护植物，同时也是栽培苹果的祖先之一。通过比较新疆野苹果不同种群的叶绿体基因组结构特征差异，来阐明新疆野苹果谱系分化格局与物种演化历史。利用Illumina NovaSeq平台对来自16个不同地区的种群进行了全基因组测序，每个种群选取了一个代表性个体。测序数据经过质量控制后，进行了基因组组装和功能注释。随后，对组装得到的基因组进行了深入的结构分析和谱系分化研究。结果显示：新疆野苹果叶绿体基因组序列全长160195~160279 bp，具有典型的四分结构。叶绿体基因组共注释到131个基因；检测出长重复序列48~58个，简单重复序列93~101个。新疆野苹果与苹果属其他物种的叶绿体基因组在IR（反向重复）区域的变异程度较低，而检测到的变异主要发生在非编码区域。系统发育方面，新疆野苹果最终分成了三个谱系，谱系Ⅰ主要分布在东部，谱系Ⅱ与谱系Ⅲ主要分布在西部。谱系Ⅰ与谱系Ⅱ的分化时间为1.74 Ma，谱系Ⅰ、谱系Ⅱ与谱系Ⅲ的分化时间为2.28 Ma。新疆野苹果遗传分化受到了第四纪气候变化的影响。相较于国外分布的新疆野苹果，中国分布的新疆野苹果遗传多样性较低，对于分布在中国的新疆野苹果需分别采取不同的方式保护，并着重关注遗传多样性较高的塔城区域。

关键词：新疆野苹果；叶绿体基因组；系统发育；分化时间；保护遗传

新疆野苹果[*Malus sieversii* (Ledeb.) M. Roem.]别名塞威氏苹果，是蔷薇科（Rosaceae）苹果属（*Malus* Mill.）植物。Velasco等[1]通过对新疆野苹果的基因组数据进行分析，发现其为栽培苹果（*M. domestica*）的祖先之一。目前，新疆野苹果仅在亚洲中部干旱区中的河谷与前山带地区片段化分布，相较于其他中亚干旱区植物，其分布区具有温暖、湿润的特征。作为第三纪孑遗物种[2]，新疆野苹果具有多种优良性状，常作为西北地区优秀的砧木资源，是苹果育种的重要种质资源，因此，新疆野苹果被认为是苹果基因库的重要组成部分[3]。但由于环境因素和人为因素的影响，新疆野苹果的种群面积正在不断地减小[4]。作为国家二级保护植物，研究新疆野苹果不仅可以了解其种质资源多样性，在此基础上解析干旱区山地孑遗落叶阔叶林植物进化历史，还可以依此划定科学遗传保护单元，提出新疆野苹果资源保护的相关建议。目前，已有研究基于核DNA小片段、SSR（简单重复序列）位点以及简化基因组SNP（单核苷酸多态性）位点对新疆野苹果种群的遗传多样性与谱系分化进行了较为系统的研究[5-7]，然而，相较于核DNA和简化基因组，母系遗传的叶绿体基因组对高等植物谱系进化历史的探讨更具优势，但利用叶绿体基因组变异对新疆野苹果不同种群进行谱系进化分析的研究鲜见报道。因此，目前亟须基于叶绿体基因组变异从母系遗传的角度深入解析新疆野苹果的谱系进化历史。

叶绿体是一种具有自主遗传系统的细胞器，是植物细胞中进行光合作用的重要场所，在高等植物

收稿日期：2024-01-04； **修订日期：**2024-10-06

基金项目：国家自然科学基金（32170391）；中国科学院战略生物资源计划（KFJ-BRP-007-008）；中国科学院"西部青年学者"项目（2022-XBQNXZ-007）

作者简介：张健（1999-），男，硕士研究生，主要从事干旱区植物保护遗传学研究. E-mail: zj13201654032@163.com

通讯作者：张宏祥. E-mail: zhanghx561@ms.xjb.ac.cn

的能量转化中发挥着重要作用;同时,它含有一套完整的遗传信息,能够自我表达和复制,是半自主型细胞器,属于单亲遗传,其中被子植物的叶绿体一般为母系遗传。与核基因组相比,植物叶绿体基因组在结构方面具有相对保守性。高等植物叶绿体基因组由于具有结构简单保守,包含大量遗传信息,碱基替换水平处于中间水平,分子进化速度在编码区和非编码区的差别很大,且为单亲遗传等特征,被广泛应用于系统发育推断和群体遗传学等研究中[8]。此外,叶绿体基因组中存在有大量的简单重复序列(Simple Sequence Repeats, SSR),其常用于遗传多样性研究,基因定位和图谱构建,品种鉴定和种质资源保存等各个方面。同时在叶绿体基因组中存在着一些单核苷酸多态性(Single Nucleotide Polymorphism, SNP)的热点区域,可产生足够的信息用于物种分类和鉴定[9],这大大增加了基于叶绿体基因组序列探讨新疆野苹果谱系进化历史研究的必要性。

目前,大量植物的叶绿体基因组由于测序技术的发展被测序、研究和报道,其中新疆野苹果的叶绿体基因组在2019年就已被报道[10],报道中的新疆野苹果叶绿体基因组仅仅进行了叶绿体基因组本身特征的描述,未与其他种的叶绿体基因组进行对比,且报道的叶绿体基因组仅覆盖了一个种群。综上所述,本研究通过对不同种群新疆野苹果的叶绿体基因组进行高通量测序及后续组装,深入分析了新疆野苹果叶绿体基因组的多个方面,包括基因结构、功能特征、重复序列、简单序列重复(SSR)位点、系统发生关系以及种群分化时间的估算。这些分析有助于揭示新疆野苹果的遗传多样性水平,阐明其谱系分化和演化历史。基于这些发现,本研究最终提出了针对新疆野苹果资源保护的策略和建议,旨在为该物种的长期保护和可持续利用提供科学依据。

1　材料与方法

1.1　实验材料、测序及叶绿体基因组组装

广泛采集了新疆野苹果的样本,覆盖了中国新疆6个种群,并扩展至邻近国家,包括吉尔吉斯斯坦、哈萨克斯坦和塔吉克斯坦,共计10个种群。从每个种群中精心选取了1个个体,并采集了其硅胶干燥的叶片样品,以便进行后续的高通量测序和基因组分析。其样品收集涵盖了新疆野苹果主要的分布区(表1)。采集样品送至上海派森诺生物科技股份有限公司,由公司进行原始数据测序,测序平台为Illumina NovaSeq平台。对下机数据使用FastQC(V 0.11.90)[11]进行质量控制后。通过Fast-Plast(V 1.2.9)对所获得的数据进行叶绿体基因组组装[12]。组装后使用PGA(V 1.3)软件进行叶绿体基因组的注释[13]。再对组装后的数据利用GeneiousPrime 2021版[14]软件进行人工矫正。组装完成的叶绿体基因组上传至NCBI数据库(获取号:PP24956–PP24971)。

表1　新疆野苹果样品采集信息

Tab. 1　Sample collection information of *Malus sieversii*

种群编号	国家来源	经度	纬度
EM	中国	84.00°E	46.36°N
GL	中国	82.86°E	43.26°N
HC	中国	80.78°E	44.43°N
JEJ2	吉尔吉斯斯坦	71.95°E	41.84°N
JEJ3	吉尔吉斯斯坦	72.96°E	41.36°N
JEJ5	吉尔吉斯斯坦	72.75°E	40.45°N
JEJ6	吉尔吉斯斯坦	73.34°E	41.21°N
NHZ	哈萨克斯坦	70.26°E	42.67°N
QMKT	哈萨克斯坦	69.95°E	41.94°N
TJK1	塔吉克斯坦	68.82°E	38.81°N
TJK2	塔吉克斯坦	69.68°E	38.72°N
TL	中国	83.54°E	46.13°N
XY	中国	83.61°E	43.38°N
YLALT	哈萨克斯坦	77.67°E	43.37°N
YN	中国	81.62°E	44.12°N
ZGR	哈萨克斯坦	80.96°E	45.64°N

1.2　叶绿体基因组结构分析与功能分析

在使用PGA软件对叶绿体进行注释后,再用Geneious Prime 2021版对新疆野苹果的叶绿体基因组进行基因组成和基因功能分析。使用OGDRAW在线软件进行新疆野苹果的叶绿体基因组物理图谱的绘制(https://chlorobox.mpimp-golm.mpg.de/OG-Draw.html)。

1.3　重复序列和SSR位点分析

利用在线软件REPuter[15](https://bibiserv.ce-bitec. uni-bielefeld)对新疆野苹果的叶绿体基因组进行重复序列分析,其中包括同向、反向、互补、与回文重复。软件参数为重复序列数目上限1000,最小重复长度30,汉明距离3,编辑距离为默认值。使用MISA[16]检测其中存在的简单重复序列,其中单核苷酸的重复次数设置为10次,二核苷酸设置为5

次，三核苷酸设置为4次。其他类型的核苷酸均设置为3次。其他参数均为默认值。

1.4 叶绿体基因组比较分析

使用CPJSdraw（V 0.0.1）[17]软件比较新疆野苹果8个近缘物种的IR边界信息。通过在线软件mVISTA[18]（https://genome.lbl.gov/vista/index.shtml）中的模型来比较新疆野苹果与上述近缘种之间叶绿体基因组序列的差异。

1.5 系统发育分析

为了阐明新疆野苹果同苹果属其他物种的系统发育关系，从GenBank数据库（https://www.ncbi.nlm.nih.gov/nuccore）中下载了苹果属的9种植物共11个叶绿体基因组，并将梨属的*Pyrus ussuriensis*与李属的*Prunus salicina*作为外类群，使用MAFFT(v 7.520)[19]先进行比对，将得到的fasta文件使用Geneious Prime 2021版转化为nex文件格式。然后使用BEAST (v1.10)[20]软件估算分化时间，并设置化石校正点：1号化石点为10 Ma，2号化石点为20 Ma，3号化石点为55 Ma[21-23]，MCMC设置链长为1000000000，其余参数均为默认值。另外，通过IQ-TREE (V2.2.2.6)[24]构建了ML系统发育树，其中IQ-TREE选择的最佳模型为TVM+F+R2。

2 结果与分析

2.1 叶绿体基因组结构、功能及特征

对采集的16个样本进行基因组结构、功能及特征分析，发现新疆野苹果叶绿体基因组全长为160195~160279 bp，平均长度为160253.56 bp。编码基因数目为131个，其中rRNA编码基因为8个，tRNA编码基因为37个，蛋白质编码基因为86个。新疆野苹果的叶绿体基因组具有经典的环状结构（图1），其

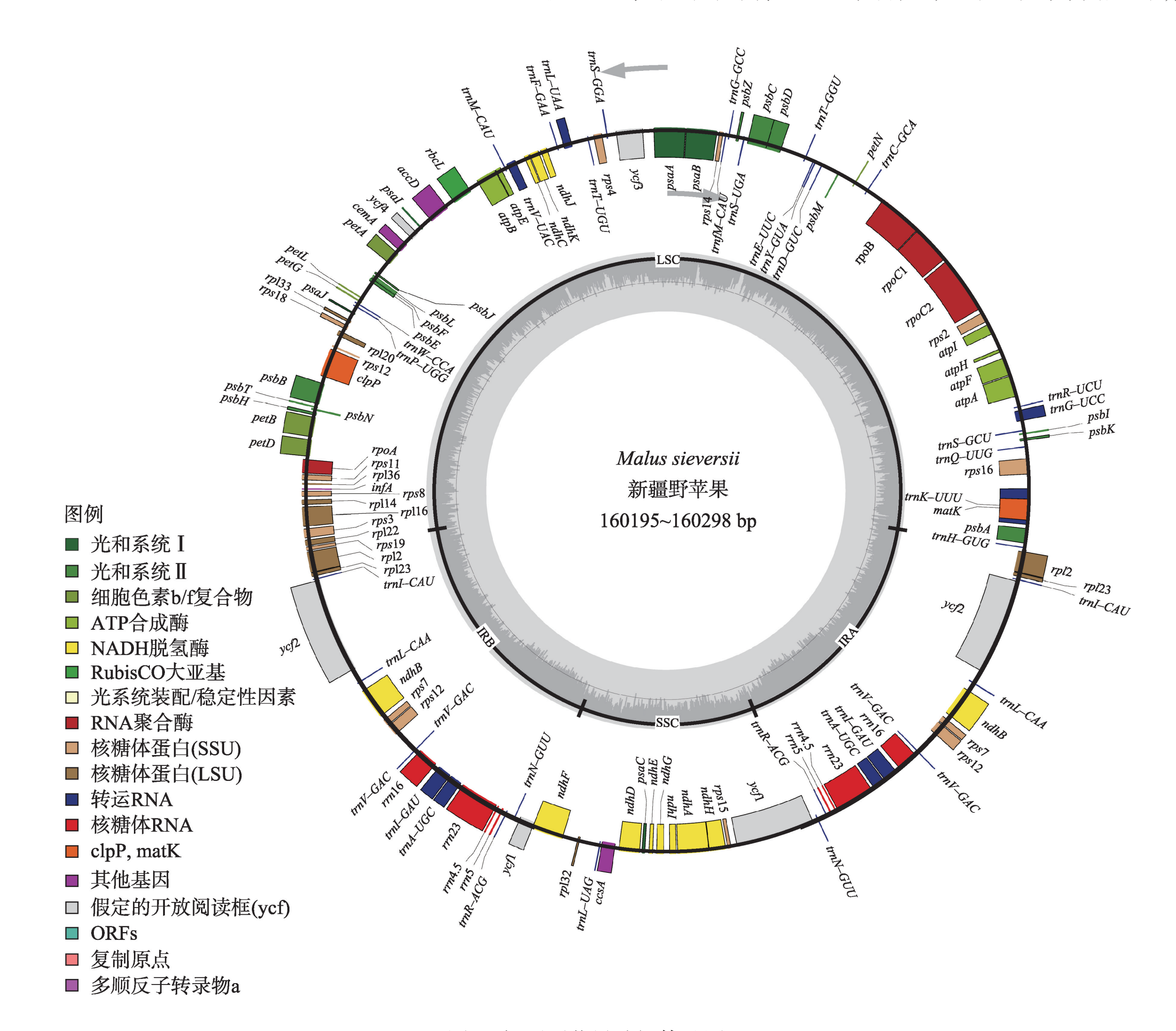


图1 新疆野苹果叶绿体基因组

Fig. 1 Chloroplast genome of *Malus sieversii*

中大单拷贝区(Large Single Copy region,LSC)的长度平均值为88352.25 bp,小单拷贝区(Small Single Copy region,SSC)的长度平均值为26360.125 bp。新疆野苹果的GC含量为35.7%。此外,新疆野苹果的叶绿体基因组中检测出具有内含子的基因为24个,其中20个基因仅具有一个内含子占绝大部分,其余的基因则具有2个内含子(表2)。

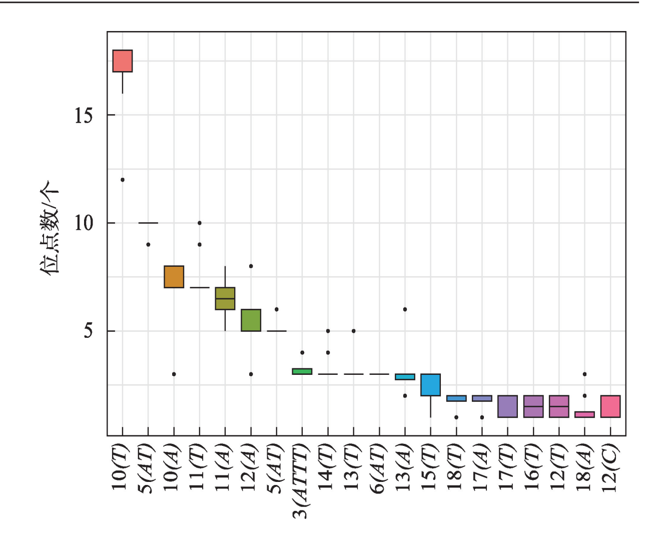
表2　新疆野苹果叶绿体基因组中含内含子的基因

Tab. 2　Genes with introns in the chloroplast genome of *Malus sieversii*

基因名称	基因数量/个	内含子数量/个	基因组
ndhA	1	1	NADH脱氢酶的亚基
ndhB	2	1	NADH脱氢酶的亚基
petB	1	1	细胞色素b/f复合体的亚基
petD	1	1	细胞色素b/f复合体的亚基
atpF	1	1	ATP合酶的亚基
*rpl*16	1	1	大核糖体亚基的蛋白质
*rpl*2	2	1	大核糖体亚基的蛋白质
*rps*12	2	2	小核糖体亚基的蛋白质
*rps*16	1	1	小核糖体亚基的蛋白质
*rpoC*1	1	1	RNA聚合酶的亚基
trnA-UGC	2	1	转运RNA
trnG-GCC	1	1	转运RNA
trnG-UCC	1	1	转运RNA
trnK-UUU	1	1	转运RNA
trnL-UAA	1	1	转运RNA
trnT-UGU	1	1	转运RNA
trnV-UAC	1	1	转运RNA
trnS-GCU	1	1	转运RNA
clpP	1	2	蛋白酶
*ycf*3	1	2	保守的假设性叶绿体开放阅读框

2.2　重复序列和SSR位点分析

在新疆野苹果叶绿体基因组中检测到48~58个30~50 bp的重复序列,其中反向重复出现次数为2~6次,回文重复出现次数为17~18次,同向重复出现次数最多,为28~34次,互补重复则出现次数最少,仅有1~3次。这些重复主要出现的区域是LSC区,最少出现的区域是在SSC区。这些重复序列大多在tRNA编码基因中被检测出,少部分在*ycf*1、*ycf*2、*ycf*3基因上被检测出来且主要位于内含子区域。在新疆野苹果SSR位点的检测中,大部分的样品被检测出具有93~101个SSR位点。并且在这些位点中单核苷酸重复出现的次数最多,五核苷酸重复出现的次数最少,未检测出六核苷酸重复(图2)。



注:横坐标括号内为基因的重复类型,数字为重复的次数。

图2　主要SSR位点分布

Fig. 2　Distribution of major SSR loci

2.3　叶绿体基因组的比较分析

通过CPJSdraw比较分析了苹果属9个物种的LSC、SSC和IR区的边界位置,结果表明这9个物种的IR区域长度较为保守(26353~26373 bp)。如图3所示,9个物种的LSC与IRb的边界在*rps*19基因的内部,扩张程度差异在4个碱基以内。SSC区与IRb和IRa区的边界分别在*ndhF*基因与*ycf*1基因之间,向IRb区和IRa区扩张的距离分别是12 bp与1074 bp,未发生变化。

如图4所示,灰色箭头表示基因方向,纵坐标表示50%~100%的一致度。将栽培苹果作为参照序列,可以看出新疆野苹果的叶绿体基因组同苹果属其他几种植物的叶绿体基因组有一定的区别。并且新疆野苹果与栽培苹果的区别明显少于其他品种,区别主要集中在非编码区,编码区则很少发生差异。此外IR区域较为保守,区别极小。

2.4　系统发育分析

如图5所示,分支上的数字为计算出的分化时间,系统发育树中大部分节点的支持率都为100%,可信度较高。通过最大似然(ML)树所得结果类似不再展示。从图中可以看出,新疆野苹果与栽培苹果和森林苹果等真苹果组物种的关系最近,与山荆子组关系稍近,与花楸苹果组的关系较远,离木衣海棠组的关系最远。新疆野苹果16个种群,可以分为3个谱系。其中谱系Ⅰ为吉尔吉斯斯坦:JEJ6,伊宁县:YN,新源县:XY,霍城县:HC,塔吉克斯坦:TJK2,巩留县:GL,额敏县:EM;谱系Ⅱ为吉尔吉斯斯坦:

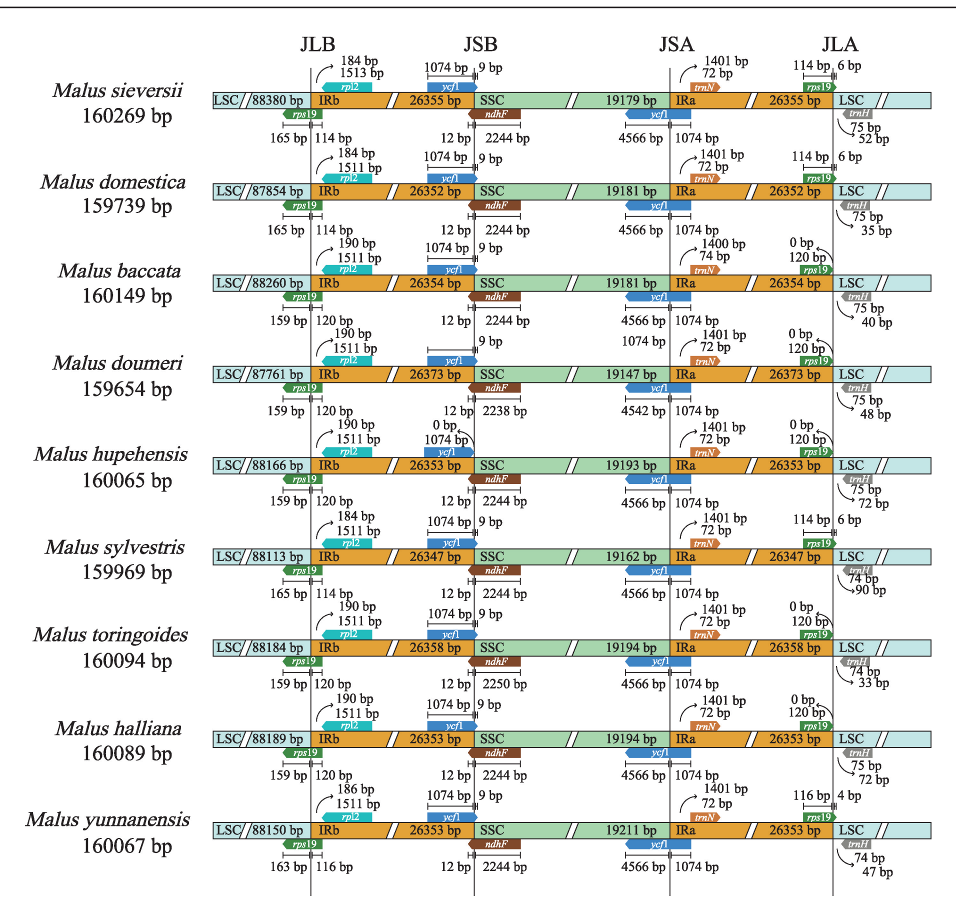
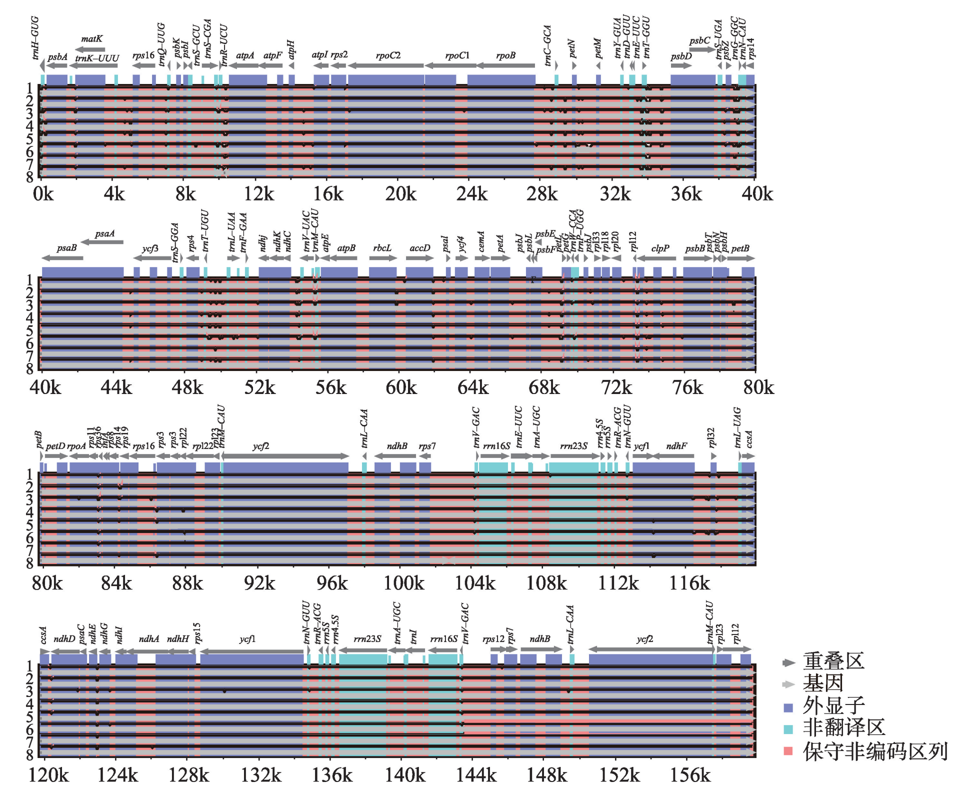


图3 苹果属9种植物叶绿体基因组边界分析

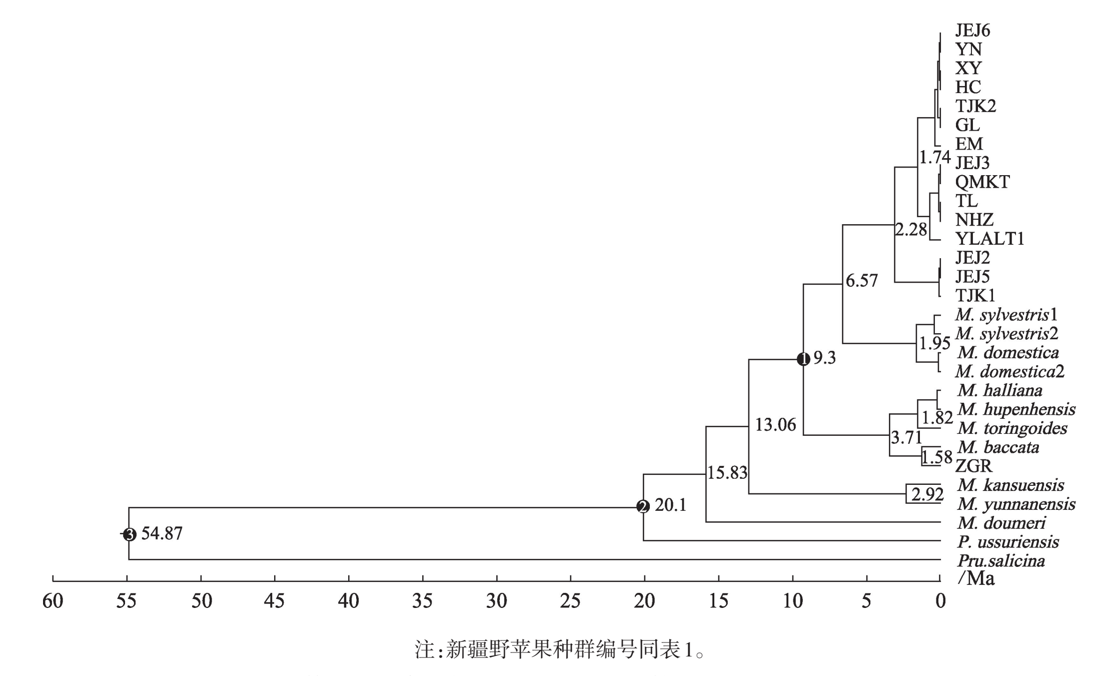
Fig. 3 Boundary analysis of chloroplast genomes of nine *Malus* species



注：1~8分别代表*M. sieversii*、*M. baccata*、*M. doumeri*、*M. halliana*、*M. hupehensis*、*M. yunnanensis*、*M. sylvestris* 和*M. toringoides*。

图4 苹果属植物叶绿体基因组序列比对

Fig. 4 A comparison of the chloroplast genome sequences of plants in the genus *Malus*



注:新疆野苹果种群编号同表1。

图5　基于叶绿体基因组序列使用BEAST软件估算的新疆野苹果谱系分化时间

Fig. 5　Lineage divergence time of *Malus sieversii* estimated based on chloroplast genome sequences using BEAST software

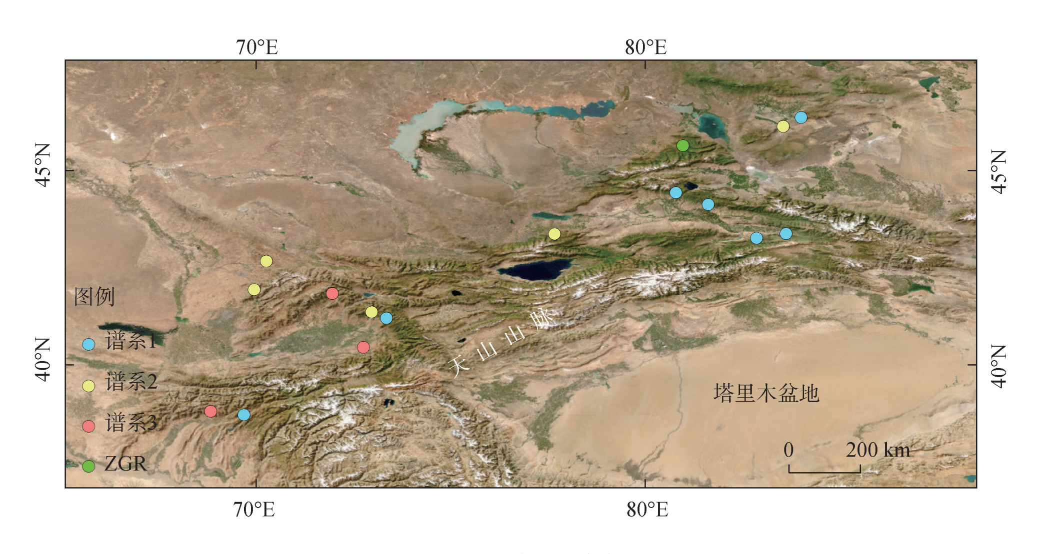


图6　居群地理分布

Fig. 6　Geographic distribution of populations

JEJ3,哈萨克斯坦:QMKT,NHZ,YLALT1,托里县:TL;谱系Ⅲ为吉尔吉斯斯坦:JEJ2,JEJ5,塔吉克斯坦:TJK1(图6)。哈萨克斯坦的ZGR则较为特殊。分化时间表明,新疆野苹果与近缘物种的分化时间发生在6.57 Ma,谱系Ⅰ与谱系Ⅱ之间的分化时间发生在1.74 Ma,谱系Ⅰ和谱系Ⅱ与谱系Ⅲ的分化时间发生在2.28 Ma。

3　讨论与结论

本研究得出,新疆野苹果叶绿体基因组序列全长为160195~160279 bp,总GC含量为36.7%,与已发表的苹果属其他物种叶绿体基因组相似[10]。经由后续的注释及分析发现,新疆野苹果的叶绿体基因组中具有131个基因。大部分为蛋白编码基因,共计86个,这与已发表的苹果属其他物种叶绿体基因数量有所差异[25],这种差异主要是由于不同研究中使用的基因组注释库不同所致。其rRNA编码基因为8个,这与大部分植物的叶绿体上的rRNA基因数目相同。tRNA的数目为37个,这也说明了新疆野苹果叶绿体基因组相对较为保守。由于苹果属的叶绿体基因组具有典型的四分体结构,因此,从IR区域边界比较便可以直观发现,IR区域变化差距

不大，并不存在IR区域丢失的状况（图3）。通过mVISTA和苹果属其他植物比对可以发现，苹果属的叶绿体基因组序列变异主要发生在非编码区，编码区的变异极少，这可能是由于编码区的基因相对保守；而IR区的编码基因几乎不存在变异，说明此区域更加保守，对叶绿体功能的正常发挥至关重要。重复序列对于分析叶绿体的基因重排、碱基替换、系统发育和基因组进化都有重要作用[8]，经研究发现，新疆野苹果的叶绿体基因组中包含了四种类型的重复序列类型，大部分为正向重复，少部分为互补序列，且大部分重复序列出现在LSC区；SSR位点出现最多的重复类型是单核苷酸重复，最少出现的重复类型是五核苷酸重复，A/T类型重复出现频率远远大于C/G类型重复，这可能是由于A/T类型更容易被解链，继而有利于基因的表达[26]。

使用叶绿体基因组进行系统发育树构建在植物分类和进化研究中具有较高的可靠性。叶绿体基因组因其保守性和单亲遗传特性，能够提供清晰的系统发育信号。此外，叶绿体基因组包含丰富的基因和非编码区信息，有助于揭示植物间的进化关系。通过BEAST分化时间分析与ML建树的结果可以得知，新疆野苹果与苹果属其他物种的系统发育关系同中国植物志中的分组一致，在系统发育关系分析的结果中可以明显看出，16个不同谱系的新疆野苹果可大致分为3个主要谱系，其中哈萨克斯坦的ZGR由于较为特殊，并未与其他谱系的新疆野苹果聚为一支，反而单独分离出来与山荆子聚为一支。Nikiforova等[22]的研究发现，新疆野苹果的叶绿体基因组与山荆子的叶绿体基因组聚为一类，同时新疆野苹果的叶绿体基因组也与垂丝海棠（*Malus halliana*）、花红（*Malus asiatica*）和海棠（*Malus adstringens*）的叶绿体基因组聚集在一起。类似的现象也在Liu等[27]的研究中被观察到，即新疆野苹果的叶绿体基因组与东方苹果（*Malus orientalis*）的叶绿体基因组聚集在一起。这些发现可能表明，苹果属植物的叶绿体基因组并非单系群，在利用叶绿体基因组构建系统发育树时可能会出现这种情况。在本研究中，样本ZGR也呈现了这一特性，因此，在分析时未对其进行深入探讨。此外，研究发现新疆野苹果与近缘物种的分化时间发生在6.57 Ma，谱系Ⅰ与谱系Ⅱ之间的分化时间发生在1.74 Ma，谱系Ⅰ和谱系Ⅱ与谱系Ⅲ的分化时间发生在2.28 Ma。从3个谱系之间的地理关系来看，谱系Ⅰ主要聚集在东部，谱系Ⅱ与谱系Ⅲ主要聚集在西部；谱系Ⅲ相较于谱系Ⅰ与谱系Ⅱ更早地分化出来，这可能说明谱系Ⅲ具有更低的遗传多样性，且未发生进一步的分化。由于第三纪晚期以来的气候干旱化和第四纪气候震荡，野果林植被呈岛屿状分布在气候相对适宜的避难所[28-29]，作为第三纪孑遗物种，新疆野苹果受到干旱化的影响，导致不同的谱系间产生了遗传分化[6]。通过对新疆野苹果的叶绿体基因组进行研究，可以看出，新疆野苹果内部已经发生了一定程度的分化。故主要在沿天山山脉附近分布的新疆野苹果因干旱化等原因产生了一定的分化，且由于地理隔离使不同谱系间的基因流减少，于是形成了现有的分布格局[6]。本研究与先前基于基因组SNP数据划分的新疆野苹果地理谱系结果有所差异[30]，这可能是由于叶绿体基因组基于母系遗传方式导致其与核基因组上的遗传结构出现了差异，特别是明显看到在叶绿体系统发育树上欧洲森林苹果的关系与栽培苹果更近，而在先前全基因组测序的分析中新疆野苹果与栽培苹果的关系更近[1]。

本文分析了不同谱系的新疆野苹果叶绿体基因组结构，并比较分析了新疆野苹果与苹果属其他植物叶绿体基因组之间的关系，同时对新疆野苹果进行了系统发育关系的分析与分化时间的估算。最终，得到了大量的重复序列及SSR位点，这些位点可用于新疆野苹果遗传多样性研究与核心种质资源筛选；此外，亦从叶绿体基因组的角度确定了新疆野苹果在苹果属中的系统发育地位，并发现了不同地理谱系间的新疆野苹果叶绿体基因组存在一定的差异；叶绿体系统发育树表明，谱系Ⅰ与谱系Ⅱ之间的分化时间为1.74 Ma，谱系Ⅰ、谱系Ⅱ与谱系Ⅲ间的分化为2.28 Ma（图5）；谱系分化时间表明，新疆野苹果遗传分化受到了第四纪气候变化的剧烈影响。然而由于自然因素与人为因素的双重影响[31-32]，近50 a来新疆野苹果原本便岌岌可危的种群面积愈发大幅下降[33]。为保护新疆野苹果核心种质资源，应该深入了解新疆野苹果的遗传多样性与谱系分化格局，对其未来的保护工作提供理论依据。目前，通过叶绿体基因组的研究可以看出，新疆地区的新疆野苹果谱系多样性明显低于国外

分布区,其中伊犁地区仅存在一个谱系,且遗传多样性较低;而塔城地区则存在2个谱系,且具有较高的遗传多样性。根据前期的研究结果[6],得知伊犁和塔城在遗传学上属于不同的单元,这意味着这两个地区在基因层面存在显著的差异。这种差异可能源于它们各自独特的自然条件、地理环境或历史演变。在自然环境,伊犁地区的降水量要显著大于塔城地区[34],这可能是导致差距的原因之一。在进行实地考察后,笔者进一步确认了这一点,并且发现塔城地区的新疆野苹果种群在数量上要少于伊犁地区。尽管种群规模较小,但塔城地区的新疆野苹果却展现出了较高的遗传多样性,这表明该地区的种群在遗传层面上具有丰富的变异和适应性。因此,为了有效地保护和维持这2个地区的新疆野苹果种群,建议采取差异化的保护措施。对于伊犁地区,可能需要关注种群的规模和分布,以及可能影响其生存的环境因素。而对于塔城地区,除了常规的保护措施外,还应该特别关注其遗传多样性的保护,因为遗传多样性是生物种群适应环境变化和生存下去的关键。对塔城地区的新疆野苹果进行详细的遗传资源调查,收集种子和植株样本,建立种质资源库,以备未来的研究和繁殖使用。还应当采取适当的措施,如优化就地保护体系等,避免其种群进一步缩小。

参考文献(References):

[1] Velasco R, Zharkikh A, Affourtit J, et al. The genome of the domesticated apple(*Malus*×*domestica* borkh.)[J]. Nature Genetics, 2010, 42(10): 833−839.

[2] 张新时. 伊犁野果林的生态地理特征和群落学问题[J]. 植物学报, 1973, 15(2): 239−253. [Zhang Xinshi. On the eco-geographical characters and the problems of classification of the wild fruit-tree forest in the Ili valley of Sinkiang[J]. Acta Botanica Sinica, 1973, 15(2): 239−253.]

[3] Volk G M, Peace C P, Henk A D, et al. DNA profiling with the 20K apple SNP array reveals *Malus domestica* hybridization and admixture in *M. sieversii*, *M. orientalis*, and *M. sylvestris* genebank accessions[J]. Frontiers in Plant Science, 2022, 13: 1015658.

[4] 米尔卡米力·麦麦提, 刘忠权, 马晓东, 等. 新疆野苹果的生存现状、问题及保护策略[J]. 广西植物, 2021, 41(12): 2100−2109. [Maimaiti Maiti, Liu Zhongquan, Ma Xiaodong, et al. Survival status, problems and conservation strategies of *Malus sieversii*[J]. Guihaia, 2021, 41(12): 2100−2109.]

[5] 张宏祥, 郑田勇. 生境片段化对新疆野苹果种群遗传结构的影响[J]. 干旱区研究, 2020, 37(3): 715−721. [Zhang Hongxiang, Zheng Tianyong. Effect of habitat fragmentation on the population genetic structure of *Malus sieversii*[J]. Arid Zone Research, 2020, 37(3): 715−721.]

[6] Zhang H X, Li X S, Wang J C, et al. Insights into the aridification history of central Asian mountains and international conservation strategy from the endangered wild apple tree[J]. Journal of Biogeography, 2021, 48(2): 332−344.

[7] Zhang H X, Zhang M L, Wang L N. Genetic structure and historical demography of *Malus sieversii* in the Ili valley and the western mountains of the Junggar Basin, Xinjiang, China[J]. Journal of Arid Land, 2015, 7(2): 264−271.

[8] 赵玉芬. 叶绿体基因组在植物学研究中的应用进展[J]. 生物学教学, 2022, 47(3): 83−85. [Zhao Yufen. Application progress of chloroplast genome in botany research[J]. Biology Teaching, 2022, 47(3): 83−85.]

[9] Vaughn J N, Chaluvadi S R, Tushar, et al. Whole plastome sequences from five ginger species facilitate marker development and define limits to barcode methodology[J]. PLoS One, 2014, 9(10): e108581.

[10] Naizaier R, Qu Z, Wu S, et al. The complete chloroplast genome of *Malus sieversii* (Rosaceae), a wild apple tree in Xinjiang, China[J]. Mitochondrial DNA Part B, 2019, 4(1): 983−984.

[11] Chen S F, Zhou Y Q, Chen Y R, et al. Fastp: An ultra-fast all-in-one fastq preprocessor[J]. Bioinformatics, 2018, 34(17): 884−890.

[12] Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform[J]. Bioinformatics, 2009, 25(14): 1754−1760.

[13] Qu X J, Moore M J, Li D Z, et al. PGA: A software package for rapid, accurate, and flexible batch annotation of plastomes[J]. Plant Methods, 2019, 15: 1−12.

[14] Geneious Prime[EB/OL]. https://www.geneious.com

[15] Kurtz S, Choudhuri J V, Ohlebusch E, et al. Reputer: The manifold applications of repeat analysis on a genomic scale[J]. Nucleic Acids Research, 2001, 29(22): 4633−4642.

[16] Beier S, Thiel T, Münch T, et al. MISA-web: A web server for microsatellite prediction[J]. Bioinformatics, 2017, 33(16): 2583−2585.

[17] Li H, Guo Q, Xu L, et al. CPJSdraw: Analysis and visualization of junction sites of chloroplast genomes[J]. PeerJ, 2023, 11: e15326.

[18] Frazer K A, Pachter L, Poliakov A, et al. VISTA: Computational tools for comparative genomics[J]. Nucleic Acids Research, 2004, 32(Suppl_2): W273−W279.

[19] Katoh K, Standley D M. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability[J]. Molecular Biology and Evolution, 2013, 30(4): 772−780.

[20] Suchard M A, Lemey P, BaeleA G, et al. Bayesian phylogenetic

and phylodynamic data integration using BEAST 1.10[J]. Virus Evolution, 2018, 4(1): vey016.

[21] Ma X, Cai Z, Liu W, et al. Identification, genealogical structure and population genetics of salleles in *Malus sieversii*, the wild ancestor of domesticated apple[J]. Heredity, 2017, 119(3): 185–196.

[22] Nikiforova S V, Cavalieri D, Velasco R, et al. Phylogenetic analysis of 47 chloroplast genomes clarifies the contribution of wild species to the domesticated apple maternal line[J]. Molecular Biology and Evolution, 2013, 30(8): 1751–1760.

[23] Xiang Y Z, Huang C H, Hu Y, et al. Evolution of Rosaceae fruit types based on nuclear phylogeny in the context of geological times and genome duplication[J]. Molecular Biology and Evolution, 2017, 34(2): 262–281.

[24] Minh B Q, Schmidt H A, Chernomor O, et al. IQ-TREE 2: New models and efficient methods for phylogenetic inference in the genomic era[J]. Molecular Biology and Evolution, 2020, 37(5): 1530–1534.

[25] Ya N L, Yan L L, Chao X, et al. The complete chloroplast genome sequence of *Malus toringo* (Rosaceae)[J]. Mitochondrial DNA Part B, 2020, 5(3): 2832–2833.

[26] Galtier N, Piganeau G, Mouchiroud D, et al. GC-content evolution in mammalian genomes: The biased gene conversion hypothesis [J]. Genetics, 2001, 159(2): 907–911.

[27] Liu B B, Ren C, Kwak M, et al. Phylogenomic conflict analyses in the apple genus *Malus* s.l. reveal widespread hybridization and allopolyploidy driving diversification, with insights into the complex biogeographic history in the Northern Hemisphere[J]. Journal of Integrative Plant Biology, 2022, 64(5): 1020–1043.

[28] 崔大方, 廖文波, 羊海军, 等. 中国伊犁天山野果林区系表征地理成分及区系发生的研究[J]. 林业科学研究, 2006, 19(5): 555–560. [Cui Dafang, Liao Wenbo, Yang Haijun, et al. Studies on the floristic composition and genesis of the wild fruit forest in Tianshan Mountains in China[J]. Forest Research, 2006, 19(5): 555–560.]

[29] 刘兴诗, 林培钧, 钟骏平. 伊犁野果林生境分析和发生探讨[J]. 干旱区研究, 1993, 10(3): 28–33. [Liu Xinshi, Lin Peijun, Zhong Junpin. An analysis and inquiry into the wild apple trees in Ili[J]. Arid Zone Research, 1993, 10(3): 28–33.]

[30] 张宏祥, 闻志彬, 王茜. 新疆野苹果种群遗传结构及其环境适应性[J]. 植物生态学报, 2022, 46(9): 1098–1108. [Zhang Hongxiang, Wen Zhibing, Wang Qian. Population genetic structure of *Malus sieversii* and environmental adaptations[J]. Chinese Journal of Plant Ecology, 2022, 46(9): 1098–1108.]

[31] 周小东, 常顺利, 王冠正, 等. 天山北坡中段雪岭云杉径向生长对气候变化的响应[J]. 干旱区研究, 2023, 40(8): 1215–1228. [Zhou Xiaodong, Chang Shunli, Wang Guanzheng, et al. Radial growth response of *Picea schrenkiana* to climate change in the middle section of the northern slope of the Tianshan Mountains[J]. Arid Zone Research, 2023, 40(8): 1215–1228.]

[32] 赵卓怡, 郝兴明. 基于Priestley-Taylor方法的中亚干旱区实际蒸散特征及归因[J]. 干旱区研究, 2023, 40(7): 1085–1093. [Zhao Zhuoyi, Hao Xingming. Actual evapotranspiration characteristics and attribution in arid Central Asia based on the Priestley-Taylor method[J]. Arid Zone Research, 2023, 40(7): 1085–1093.]

[33] 褚佳瑶, 冯琳骄, 侯毅兴, 等. 新疆野苹果种群受损现状[J]. 经济林研究, 2022, 40(1): 265–273. [Chu Jiayao, Feng Lingjiao, Hou Yixing, et al. Analysis on population damage of *Malus sieversii*[J]. Non-wood Forest Research, 2022, 40(1): 265–273.]

[34] 董翰林, 王文婷, 谢云, 等. 新疆气候干湿变化特征及其影响因素[J]. 干旱区研究, 2023, 40(12): 1875–1884. [Dong Hanlin, Wang Wenting, Xie Yun, et al. Climate dry-wet conditions, changes, and their driving factors in Xinjiang[J]. Arid Zone Research, 2023, 40(12): 1875–1884.]

Characteristics and evolutionary history of the chloroplast genome in *Malus sieversii*

ZHANG Jian[1,2], ZHANG Hongxiang[1,3,4]

(1. State Key Laboratory of Desert and Oasis Ecology, Key Laboratory of Ecological Safety and Sustainable Development in Arid Lands, Xinjiang Institute of Ecology and Geography, Chinese Academy of Sciences, Urumqi 830011, Xinjiang, China; 2. University of Chinese Academy of Science, Beijing 100093, China; 3. Xinjiang Key Lab of Conservation and Utilization of Gene Resources, Urumqi 830011, Xinjiang, China; 4. Specimen Museum of Xinjiang Institute of Ecology and Geography, Chinese Academy of Sciences, Urumqi 830011, Xinjiang, China)

Abstract: *Malus sieversii*, a state-protected species and the progenitor of cultivated apples, is an important germplasm resource within the genus *Malus*. In this study, we aimed to compare the structural characteristics of chloroplast genomes across various populations of *M. sieversii*, clarify the lineage divergence pattern, and trace the evolutionary history of this species. We used the Illumina NovaSeq platform to conduct whole-genome sequencing of individuals from 16 different populations, with one sample representing each population. After conducting quality control on the sequencing data, we conducted genome assembly and functional annotation. Subsequently, we conducted a comprehensive structural analysis and lineage differentiation studies on the assembled genomes. The chloroplast genome length in *M. sieversii* ranged from 160195 to 160279 base pairs (bp), exhibiting a typical tetrad structure. In total, 131 genes were identified within the chloroplast genome, along with 48–58 long repeats and 93–101 simple sequence repeats. Notably, variations in the IR region between *M. sieversii* and other species in the genus were minimal, predominantly occurring in noncoding regions. Phylogenetic analysis revealed that *M. sieversii* clusters into three distinct lineages: lineage I, primarily occupying the eastern part of the distribution range, and lineages II and III, predominantly found in the west. The divergence time between lineages I and II was approximately 1.74 million years ago (Ma), while the divergence between lineages I, II, and III was around 2.28 Ma. These findings indicate that the lineage divergences of *M. sieversii* were significantly influenced by climate changes during the Quaternary period. Compared to internationally distributed populations, *M. sieversii* in China shows relatively low genetic diversity. Therefore, tailored conservation strategies should be implemented for *M. sieversii* across different regions, with particular emphasis on protecting genetically diverse populations in the Tacheng area.

Keywords: *Malus sieversii*; chloroplast genome; phylogeny; divergence time; conservation genetics